Stochastic Processes and their Applications 35 (1990) 141-148
North-Holland

TIME INTEGRATED LEAST SQUARES ESTIMATORS OF REGRESSION PARAMETERS OF INDEPENDENT STOCHASTIC PROCESSES

Tiee-Jian WU*
Department of Mathematics, University of Houston, Houston, TX 77004, USA

M.T. WASAN
Department of Mathematics and Statistics, Queen's University, Kingston, Ont., K7L 3N6 Canada

Received 1 November 1988
Revised 26 June 1989

Industrial processes may be continuously monitored by instruments under control of microprocessors. Thus the data are usually obtained in the form of sets of (continuous) curves over certain time intervals. This paper presents a method of estimating regression parameters in terms of the sample paths from independent stochastic processes. Time integrated least squares estimators of the parameters are obtained which are unbiased, translation invariant, consistent and asymptotically jointly normal. Since technically it is difficult to compute these estimators, using analog-to-digital conversion of continuous processes which are time sampled at regular intervals, optimal approximations of the estimators are considered which are very easily computable and their asymptotic properties are appended.

time integrated process * regression parameters * sample paths of stochastic processes

Introduction

Industrial processes may be continuously monitored by instruments under control of microprocessors. Thus the data are usually obtained in the form of sets of (continuous) curves over certain time intervals (i.e., the sample paths of the stochastic processes). This paper presents a method of estimating regression parameters in terms of the sample paths from independent stochastic processes. In Section 2 time integrated least squares estimators of the parameters are obtained. In Theorem 1 they are shown to be unbiased, translation invariant, consistent in mean square sense, and asymptotically jointly normal.

But in practice it is difficult to compute any functionals based on these continuous (at least in time) sample paths. Thus one has to resort to analog-to-digital conversion of continuous processes which are usually time sampled at regular intervals. In

* The research was done while this author was visiting Queen's University, Kingston, Ontario, Canada supported by NSERC grant OGP 0003117.

Section 3 optimal approximations of the regression estimators are obtained which are easily computable. In Theorem 2 and Corollary 1 they are shown to be asymptotically unbiased, L_2-consistent, and asymptotically jointly normal. In Section 4, estimation of the drift parameter is considered.

1. Mathematical models

Let $Y_1(t), \ldots, Y_N(t)$, $0 \leq t \leq 1$, be N (≥ 2) independent stochastic processes such that for $i = 1, \ldots, N$,

$$Y_i(t) = (\alpha + \beta x_i) g(t) + \varepsilon_i(t), \quad 0 \leq t \leq 1, \tag{1}$$

where α and β are unknown real parameters, x_i's are known real regression constants that are not all equal, g is a known real-valued measurable non-random function satisfying

$$0 < \|g\|_2 < \infty \tag{2}$$

(we denote $\|\cdot\|_2 = \{\int_0^1 (\cdot(t))^2 \, dt\}^{1/2}$), and $\varepsilon_i(t)$, $i = 1, \ldots, N$, is a process with stationary and independent increments and satisfies

$$\varepsilon_i(0) = 0, \qquad \|\varepsilon_i\|_2 < \infty \quad \text{with probability 1}, \tag{3}$$

$$E[\varepsilon_i(t)] = at, \qquad \operatorname{Var}[\varepsilon_i(t)] = bt, \tag{4}$$

$$\operatorname{Cov}(\varepsilon_i(s), \varepsilon_i(t)) = b \min(s, t), \quad 0 \leq s, t \leq 1, \tag{5}$$

where a and b are known real numbers with $b > 0$.

We aim at estimating α and β based on the sample paths of $Y_1(t), \ldots, Y_N(t)$. Note that if $\varepsilon_i(t)$ is the standard Brownian motion process (resp., Poisson process, inverse Gaussian process; cf. Wasan, 1968), then (3) is satisfied, and (4)-(5) are satisfied with $a = 0$ and $b = 1$ (resp., $a = b > 0$, $a = 1$). For simplicity and without loss of generality we assume that $a = 0$ and $b = 1$ in (4)-(5) (otherwise, we could consider $b^{-1/2}(Y_i(t) - at))$ for the rest of the paper.

2. Time integrated least squares estimators of α and β

For each fixed t let

$$L(\alpha, \beta, t) = \sum_{i=1}^{N} [Y_i(t) - (\alpha + \beta x_i) g(t)]^2.$$

The classical LSE's of α and β are obtained by minimizing $L(\alpha, \beta, t)$. In our present situation, it is natural to consider the average (in time) error sum of squares $L(\alpha, \beta) = \int_0^1 L(\alpha, \beta, t) \, dt$ and choose $\hat{\alpha}$ and $\hat{\beta}$ such that $L(\hat{\alpha}, \hat{\beta}) = \min_{\alpha,\beta} L(\alpha, \beta)$.

$\hat{\alpha}$ and $\hat{\beta}$ will be called TILSE's (time integrated least squares estimators) of α and β. Solving the normal equations

$$\frac{\partial L(\alpha, \beta)}{\partial \alpha} = 0, \qquad \frac{\partial L(\alpha, \beta)}{\partial \beta} = 0, \tag{6}$$

we obtain (noting that (2)-(3) ensure $(\partial/\partial\alpha)\int_0^1 = \int_0^1 (\partial/\partial\alpha)$ and $(\partial/\partial\beta)\int_0^1 = \int_0^1 (\partial/\partial\beta)$),

$$\hat{\beta} = \|g\|_2^{-2}\left(\sum_{i=1}^N (x_i - \bar{x})^2\right)^{-1} \sum_{i=1}^N (x_i - \bar{x}) \int_0^1 Y_i(t) g(t)\,\mathrm{d}t, \tag{7}$$

$$\begin{aligned}\hat{\alpha} &= \left\{ N^{-1}\|g\|_2^{-2} \sum_{i=1}^N \int_0^1 Y_i(t) g(t)\,\mathrm{d}t \right\} - \bar{x}\hat{\beta} \\ &= N^{-1}\|g\|_2^{-2}\left(\sum_{i=1}^N (x_i - \bar{x})^2\right)^{-1} \sum_{i=1}^N \left(\sum_{j=1}^N x_j^2 - N\bar{x}x_i\right) \int_0^1 Y_i(t) g(t)\,\mathrm{d}t,\end{aligned} \tag{8}$$

where $\bar{x} = N^{-1}\sum_{i=1}^N x_i$. The estimators (7)-(8) possess some nice properties, as shown by the following result.

Theorem 1. *Under* (1)-(5), *we have*:

(i) $E(\hat{\alpha}) = \alpha$ *and* $E(\hat{\beta}) = \beta$, *all* α *and* β.

(ii) *Both* $\hat{\alpha}$ *and* $\hat{\beta}$ *are translation invariant in the sense that for any known real numbers c and d*,

$$\hat{\beta}(\boldsymbol{Y} + (\boldsymbol{c} + d\boldsymbol{x})g) = \hat{\beta}(\boldsymbol{Y}) + d, \tag{9}$$

$$\hat{\alpha}(\boldsymbol{Y} + (\boldsymbol{c} + d\boldsymbol{x})g) = \hat{\alpha}(\boldsymbol{Y}) + c, \tag{10}$$

where $\hat{\theta}(\boldsymbol{Y} + (\boldsymbol{c} + d\boldsymbol{x})g)$ *denotes the TILSE of* θ, $\theta = \alpha$ *or* β, *based on the processes* $Y_i(t) + (c + dx_i)g(t)$, $i = 1, \ldots, N$.

(iii) $$\operatorname{Var}(\hat{\alpha}) = N^{-1}\|g\|_2^{-4}\left(\sum_{i=1}^N (x_i - \bar{x})^2\right)^{-1} \sum_{i=1}^N x_i^2 \sigma_G^2,$$

$$\operatorname{Var}(\hat{\beta}) = \|g\|_2^{-4}\left(\sum_{i=1}^N (x_i - \bar{x})^2\right)^{-1} \sigma_G^2,$$

where $\sigma_G^2 = (G(1) - \bar{G})^2 + \|G - \bar{G}\|_2^2$ *with* $G(t) = \int_0^t g(s)\,\mathrm{d}s$, $0 \leq t \leq 1$, *and* $\bar{G} = \int_0^1 G(t)\,\mathrm{d}t$. *Hence* $\hat{\alpha}$ *and* $\hat{\beta}$ *are* L_2*-consistent of* α *and* β, *respectively, provided that*

$$\lim_{N\to\infty} \sum_{i=1}^N (x_i - \bar{x})^2 = \infty \quad \text{and} \quad \lim_{N\to\infty} N^{-1}\left(\sum_{i=1}^N (x_i - \bar{x})^2\right)^{-1} \sum_{i=1}^N x_i^2 = 0. \tag{11}$$

(iv) *Assume that there exist* $0 < \lambda < \infty$ *and* $-\infty < \mu < \infty$ *such that as* $N \to \infty$,

$$\begin{aligned}&\max_{1\leq i\leq N} (x_i - \bar{x})^2 \left(\sum_{i=1}^N (x_i - \bar{x})^2\right)^{-1} \to 0, \\ &N^{-1}\sum_{i=1}^N (x_i - \bar{x})^2 \to \lambda^2, \qquad \bar{x} \to \mu.\end{aligned} \tag{12}$$

Then $N^{1/2}(\hat{\alpha}-\alpha, \hat{\beta}-\beta)$ *is asymptotically jointly normal* $((0,0), \|g\|_2^{-4}\sigma_G^2\Sigma)$, *where*

$$\Sigma = \begin{pmatrix} (\lambda^2+\mu^2)/\lambda^2 & -\mu/\lambda^2 \\ -\mu/\lambda^2 & 1/\lambda^2 \end{pmatrix}.$$

Proof. By (2)-(5) (recall that we assumed $a=0$ and $b=1$) and using Fubini's theorem and integration by parts, we have

$$E\int_0^1 \varepsilon_i(t)g(t)\,\mathrm{d}t = \int_0^1 E\varepsilon_i(t)g(t)\,\mathrm{d}t = 0, \tag{13}$$

$$\begin{aligned} &E\left(\int_0^1 \varepsilon_i(t)g(t)\,\mathrm{d}t\right)^2 \\ &= E\left\{\int_0^1 g(t)\left(\int_0^t \varepsilon_i(t)\varepsilon_i(s)g(s)\,\mathrm{d}s + \int_t^1 \varepsilon_i(t)\varepsilon_i(s)g(s)\,\mathrm{d}s\right)\mathrm{d}t\right\} \\ &= \int_0^1 g(t)\int_0^t sg(s)\,\mathrm{d}s\,\mathrm{d}t + \int_0^1 tg(t)(G(1)-G(t))\,\mathrm{d}t \\ &= (G(1)-\bar{G})^2 + \|G-\bar{G}\|_2^2 = \sigma_G^2. \end{aligned} \tag{14}$$

From (1), (7)-(8), and (13)-(14) we conclude (i) and (iii) quickly. (9) follows from (7), and (10) follows from (8) and (9). This proves (ii).

We now prove (iv). Put $\hat{\delta} = \hat{\alpha} + \bar{x}\hat{\beta}$. Then $\mathrm{Var}(N^{1/2}\hat{\delta}) = \|g\|_2^{-4}\sigma_G^2$ and $E\hat{\delta} = \alpha + \bar{x}\beta$ by (i). By (iii) and (12), we have $\lim_{N\to\infty}\mathrm{Var}(N^{1/2}\hat{\beta}) = \|g\|_2^{-4}\sigma_G^2\lambda^{-2}$. Simple computations indicate that $\mathrm{Cov}(\hat{\delta}, \hat{\beta}) = 0$ for all $N \geqslant 2$. In view of Lemma 5.1 of Adichie (1967), (iv) will be proved if we show that $N^{1/2}(\hat{\delta}-E\hat{\delta}, \hat{\beta}-\beta)$ is asymptotically jointly normal $((0,0), \|g\|_2^{-4}\sigma_G^2\Delta)$, where

$$\Delta = \begin{pmatrix} 1 & 0 \\ 0 & 1/\lambda^2 \end{pmatrix}.$$

But this is implied by the fact that for any real a_1 and a_2, $a_1N^{1/2}(\hat{\delta}-E\hat{\delta}) + a_2N^{1/2}(\hat{\beta}-\beta)$ is asymptotically normal $(0, (a_1^2+a_2^2\lambda^{-2})\|g\|_2^{-4}\sigma_G^2)$, which can be seen from (12) and from Theorem V.1.2 of Hájek and Sĭdák (1967). □

3. Optimal approximations of $\hat{\alpha}$ and $\hat{\beta}$ when $g(t)=t$

In practice it is sometimes very difficult (e.g., when $\varepsilon_i(t)$ is the standard Brownian motion process) to compute the realization of $\hat{\alpha}$ and $\hat{\beta}$ due to the difficulty in computing $\int_0^1 Y_i(t)g(t)\,\mathrm{d}t$. Hence, we need to approximate $\hat{\alpha}$ and $\hat{\beta}$ via approximating the preceding integral by another suitable r.v. whose realization is easily computable. Quite often one uses a sum of equally spaced rectangular areas to approximate the areas under a curve. Thus we shall consider r.v.'s of forms $\sum_{j=1}^{m} cY_i(jc)g(jc)$, $m \geqslant 2$, where $0 < c \leqslant m^{-1}$. We aim at choosing c such that

$$R_c = E\left\{\int_0^1 \varepsilon_i(t)g(t)\,\mathrm{d}t - \sum_{j=1}^{m} c\varepsilon_i(jc)g(jc)\right\}^2 \tag{15}$$

is minimized. Note that R_c is independent of i.

By (2)-(5), Fubini's theorem and integration by parts, we get

$$E\left(\varepsilon_i(jc)\int_0^1 \varepsilon_i(t)g(t)\,\mathrm{d}t\right) = \int_0^{jc} tg(t)\,\mathrm{d}t + jc(G(1) - G(jc))$$

$$= jcG(1) - \int_0^{jc} G(t)\,\mathrm{d}t,$$

which, together with (14), yields

$$R_c = \sigma_G^2 - 2c\sum_{j=1}^{m} g(jc)\left(jcG(1) - \int_0^{jc} G(t)\,\mathrm{d}t\right)$$

$$+ c^3\sum_{j=1}^{m} g^2(jc)j + 2c^3 \sum\sum_{1\leqslant i<j\leqslant m} g(ic)g(jc)i. \tag{16}$$

From (16), it is clear that one hardly can solve the problem of minimizing R_c without knowing the exact form of g. For the rest of this section, we shall confine ourselves to the case

$$g(t) = t, \quad 0 \leqslant t \leqslant 1, \tag{17}$$

because it is of practical interest and importance.

Lemma 1. *Assume* (3)-(5) *and* (17). *Then* R_c *is minimized by choosing*

$$c^* = 3[(3m+1)(3m+2)]^{-1/2}. \tag{18}$$

Furthermore,

$$R^* = \min R_c = \tfrac{1}{15}\{2 - 27m(m+1)(2m+1)[(3m+1)(3m+2)]^{-3/2}\}$$

($= \mathrm{o}(1)$ *as* $m \to \infty$).

Proof. For (17) we have $G(t) = \frac{1}{2}t^2$, $\bar{G} = \frac{1}{6}$ and $\sigma_G^2 = \frac{2}{15}$. It follows from (16) that

$$R_c = \tfrac{2}{15} - \tfrac{1}{6}c^3 m(m+1)(2m+1) + \tfrac{1}{90}c^5 m(m+1)(2m+1)(3m+1)(3m+2).$$

Now setting $\mathrm{d}R_c/\mathrm{d}c = 0$ and solving the equation for c, we obtain (18). The rest of the proof is straightforward. □

Let us define for each $0<c\leqslant m^{-1}$,

$$\tilde{\beta}_m(c)=3\left(\sum_{i=1}^{N}(x_i-\bar{x})^2\right)^{-1}\sum_{i=1}^{N}(x_i-\bar{x})c^2\sum_{j=1}^{m}jY_i(jc) \tag{19}$$

and

$$\begin{aligned}\tilde{\alpha}_m(c)&=\left\{3N^{-1}\sum_{i=1}^{N}c^2\sum_{j=1}^{m}jY_i(jc)\right\}-\bar{x}\tilde{\beta}_m(c)\\&=3N^{-1}\left(\sum_{i=1}^{N}(x_i-\bar{x})^2\right)^{-1}\sum_{i=1}^{N}\left(\sum_{k=1}^{N}x_k^2-Nx_i\bar{x}\right)c^2\sum_{j=1}^{m}jY_i(jc).\end{aligned} \tag{20}$$

Theorem 2. *Assume* (1), (3)–(5) *and* (17). *Then for each* $N\geqslant 2$ *and* $m\geqslant 2$:

(i) $$\min_{0<c\leqslant 1/m}\operatorname{Var}(\hat{\beta}-\tilde{\beta}_m(c))=\operatorname{Var}(\hat{\beta}-\tilde{\beta}_m(c^*))=9\left(\sum_{i=1}^{N}(x_i-\bar{x})^2\right)^{-1}R^*,$$

$$\begin{aligned}\min_{0<c\leqslant 1/m}\operatorname{Var}(\hat{\alpha}-\tilde{\alpha}_m(c))&=\operatorname{Var}(\hat{\alpha}-\tilde{\alpha}_m(c^*))\\&=9N^{-1}\left(\sum_{i=1}^{N}(x_i-\bar{x})^2\right)^{-1}\sum_{i=1}^{N}x_i^2R^*.\end{aligned}$$

(ii) $$E\tilde{\beta}_m(c^*)=2^{-1}\beta m(m+1)(2m+1)(c^*)^3,$$

$$E\tilde{\alpha}_m(c^*)=2^{-1}\alpha m(m+1)(2m+1)(c^*)^3.$$

(iii) $$\operatorname{Var}\tilde{\beta}_m(c^*)=3(c^*)^5m(m+1)(2m+1)(2m^2+2m+1)\Big/\left[10\sum_{i=1}^{N}(x_i-\bar{x})^2\right],$$

$$\begin{aligned}\operatorname{Var}\tilde{\alpha}_m(c^*)&=3(c^*)^5m(m+1)(2m+1)(2m^2+2m+1)\\&\quad\times\sum_{i=1}^{N}x_i^2\Big/\left[10N\sum_{i=1}^{N}(x_i-\bar{x})^2\right],\end{aligned}$$

where c^* *and* R^* *are defined in Lemma* 1.

Proof. The proof follows from Lemma 1 and routine computations. □

We now consider the case $m=rN$ for some $r>0$. From Theorems 1-2 we have:

Corollary 1. *Assume* (1), (3)–(5) *and* (17). *Then for each* $r>0$ *it holds as* $N\to\infty$:

(i) $\tilde{\alpha}_{rN}(c^*)[\tilde{\beta}_{rN}(c^*)]$ *is asymptotically unbiased for* $\alpha[\beta]$.

(ii) $\tilde{\alpha}_{rN}(c^*)[\tilde{\beta}_{rN}(c^*)]$ *is* L_2*-consistent for* $\alpha[\beta]$ *under* (11).

(iii) $(rN)^{1/2}(\tilde{\alpha}_{rN}(c^*)-\alpha,\ \tilde{\beta}_{rN}(c^*)-\beta)$ *is asymptotically jointly normal* $((0,0),\frac{6}{5}r\Sigma)$ *under* (12), *where* Σ *is given in Theorem* 1(iv).

Proof. (i)–(ii) are immediate. We now prove (iii). By (12) and Theorem 2(i)–(ii), we have as $N \to \infty$,

$$rNE(\tilde{\alpha}_{rN}(c^*)-\hat{\alpha})^2 \to 0, \qquad rNE(\tilde{\beta}_{rN}(c^*)-\hat{\beta})^2 \to 0,$$

which in turn implies that for any real a_1 and a_2, $a_1(rN)^{1/2}(\tilde{\alpha}_{rN}(c^*)-\alpha)+a_2(rN)^{1/2}(\tilde{\beta}_{rN}(c^*)-\beta)$ has the same limiting distribution as $a_1(rN)^{1/2}(\hat{\alpha}-\alpha)+a_2(rN)^{1/2}(\hat{\beta}-\beta)$. The proof follows from Theorem 1(iv) immediately. □

4. Example: Estimation of the drift parameter

In this section we present an example to illustrate the theorems of Sections 2–3. We concentrate on the case $\beta = 0$ in (1), i.e.,

$$Y_i(t) = \alpha g(t) + \varepsilon_i(t), \quad 0 \leq t \leq 1, \tag{21}$$

with $g(t)$ and $\varepsilon_i(t)$ satisfying conditions (2) through (5).

Arguing analogously as in deriving (7)–(8), we obtain

$$\hat{\alpha} = N^{-1}\|g\|_2^{-2} \sum_{i=1}^{N} \int_0^1 Y_i(t) g(t)\, \mathrm{d}t, \tag{22}$$

which is the TILSE of α for the model (21). From (13)–(14) we easily get $E\hat{\alpha} = \alpha$ for all α, $\operatorname{Var} \hat{\alpha} = N^{-1}\|g\|_2^{-4}\sigma_G^2$, $\hat{\alpha}$ is translation invariant in the sense specified by (10) (with $d = 0$), and $\hat{\alpha}$ is asymptotically normal $(\alpha, \operatorname{Var} \hat{\alpha})$.

We now consider the case (17) (if $\varepsilon_i(t)$ is the standard Brownian motion process, then α is the drift parameter for the Brownian motion process $Y_i(t)$). For this case, by Lemma 1 and Theorem 2, the optimal approximation of (22) is given by

$$\tilde{\alpha}_m(c^*) = 3N^{-1} \sum_{i=1}^{N} (c^*)^2 \sum_{j=1}^{m} jY_i(jc^*) \tag{23}$$

with

$$\operatorname{Var}(\hat{\alpha} - \tilde{\alpha}_m(c^*)) = 9N^{-1}R^*,$$

$$E\tilde{\alpha}_m(c^*) = 2^{-1}\alpha m(m+1)(2m+1)(c^*)^3$$

and

$$\operatorname{Var} \tilde{\alpha}_m(c^*) = 3(c^*)^5 m(m+1)(2m+1)(2m^2+2m+1)/(10N)$$

(c^* and R^* are given in Lemma 1). Furthermore, (23) possesses all the nice asymptotic properties stated in Corollary 1 (in our present situation, conditions (11)–(12) are dropped, and $(rN)^{1/2}(\tilde{\alpha}_{rN}(c^*)-\alpha)$ is asymptotically normal $(0, \frac{6}{5}r)$).

Remark. From Theorem 1(iv) and Corollary 1(iii), one could easily formulate asymptotic confidence intervals for α and β.

Acknowledgement

We would like to express our sincere thanks to the referee for comments which improved the presentation of this paper.

References

J.N. Adichie, Estimates of regression parameters based on rank tests, Ann. Statist. 38 (1967) 894–904.

J. Hájek and Z. Sǐdák, Theory of Rank Tests (Academic Press, New York, 1967).

M.T. Wasan, On an inverse Gaussian process, Skand. Aktuariet. Argang LI (1968) 69–96.